**Information Retrieval**

**16 July 2019**

**Ex 1 [ranks 5]** Let us given a graph G of directed edges {(1,3), (3,1), (1,4), (2,1)}. Simulate the execution of **one step** of the PageRank algorithm, starting with all nodes having PR=1 (unnormalized), and assuming a teleportation step which favors node 3 (hence, it jumps only to it).

**Ex 2 [points 5]** Simulate the execution of the Consistent Hashing technique over the urls\_ID = {3, 4, 9, 2, 5, 7, 12, 11} and the crawlers\_ID = {1,2,3}, by defining proper hash functions in the co-domain of size m=13.

**Ex 3 [ranks 4+3]** Let us given a set of strings S = { pitom, dad, daddy, zoom }.

* Build a 2-gram index over S
* Given P = atom, show how your index executes the 1-edit error search in S.

**Ex 4 [points 5]** Show the Rocchio’s formula and comment on its application and its pros/cons.

**Ex 5 [points 4+4]** Given the two adjacency lists of a Web graph for the nodes 14 and 15, namely

14 🡪 3, 10, 11, 13, 14, 17, 19, 21, 25

15 🡪 5, 10, 11, 12, 14, 17, 19, 20, 21, 24, 33

show how the list of “15” can be compressed given the list of “14” by means of the algorithm adopted in WebGraph via (1) first copy-lists and (2) then copy-blocks.